

ePub^{WU} Institutional Repository

Sylvia Frühwirth-Schnatter

Fully Bayesian Analysis of Switching Gaussian State Space Models

Working Paper

Original Citation:

Frühwirth-Schnatter, Sylvia (2000) Fully Bayesian Analysis of Switching Gaussian State Space Models. *Forschungsberichte / Institut für Statistik*, 73. Department of Statistics and Mathematics, WU Vienna University of Economics and Business, Vienna.

This version is available at: <http://epub.wu.ac.at/812/>

Available in ePub^{WU}: July 2006

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

Fully Bayesian Analysis of Switching Gaussian State Space Models



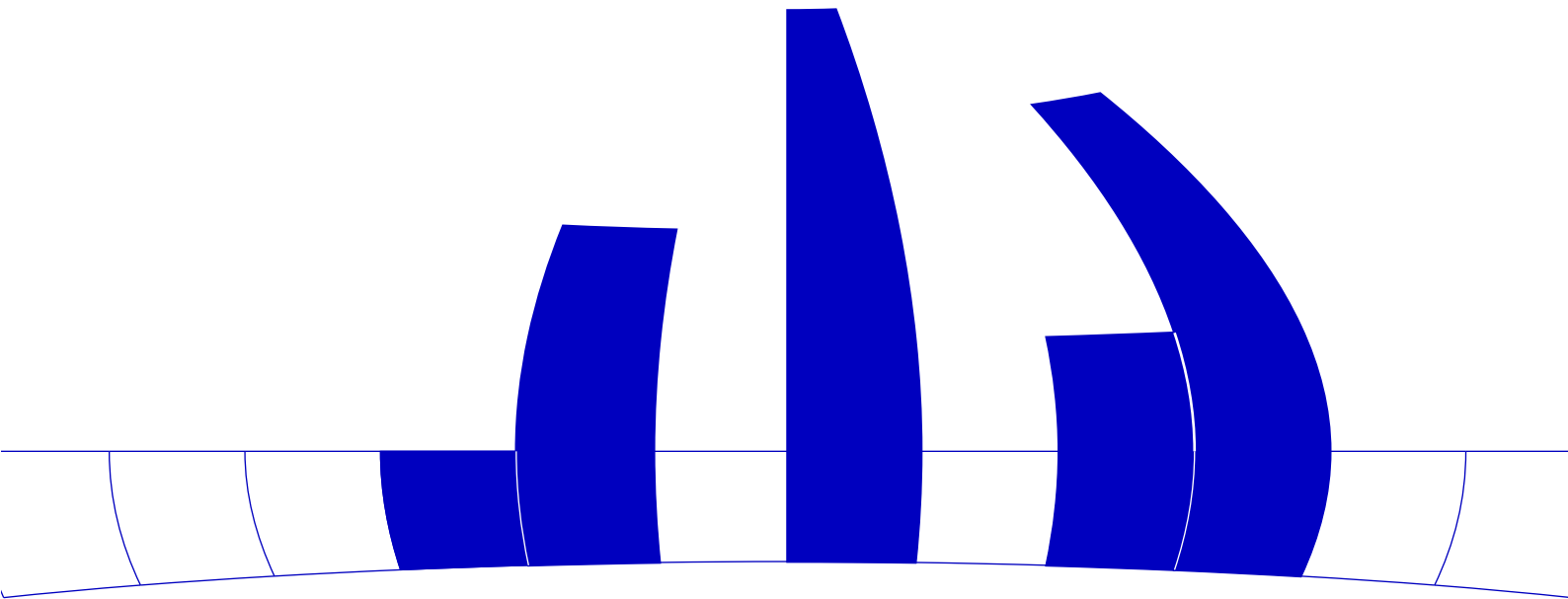
Sylvia Frühwirth-Schnatter

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 73
April 2000

<http://statmath.wu-wien.ac.at/>



2. 1. 8

Fully Bayesian Analysis of
Switching Gaussian State Space Models

Sylvia Frühwirth-Schnatter
Research Memorandum No. 73

Fully Bayesian Analysis of Switching Gaussian State Space Models

Sylvia Frühwirth-Schnatter *

April 18, 2000

Abstract

In the present paper we study switching state space models from a Bayesian point of view. For estimation, the model is reformulated as a hierarchical model. We discuss various MCMC methods for Bayesian estimation, among them unconstrained Gibbs sampling, constrained sampling and permutation sampling. We address in detail the problem of unidentifiability, and discuss potential information available from an unidentified model. Furthermore the paper discusses issues in model selection such as selecting the number of states or testing for the presence of Markov switching heterogeneity. The model likelihoods of all possible hypotheses are estimated by using the method of bridge sampling. We conclude the paper with applications to simulated data as well as to modelling the U.S./U.K. real exchange rate.

Keywords: Bayesian analysis, bridge sampling, Markov switching models, MCMC methods, model selection, state space models

1 Introduction

State space models are a well-studied tool to analyze time series $y^N = (y_1, \dots, y_N)$, where the distribution of y_t depends on a latent continuous state process $x^N = (x_0, \dots, x_N)$. A switching state space model is obtained, if we assume that additionally to the latent, continuous process x^N a discrete latent switching variable I_t taking values in $\{1, \dots, K\}$ influences the distribution of y_t . Such models have been studied by various authors (e.g. Harrison and Stevens, 1976; Shumway and Stoffer, 1991; Kim, 1993, 1994; Carter and Kohn, 1994, 1996; Shephard, 1994). Kim and Nelson (1999) present an excellent review of the current state of art.

A typical way of including a switching mechanism into a Gaussian state space model is to assume that one of the variances, e.g. the variance appearing in the observation equation, is heteroskedastic and switches between various values $\theta_1^I, \dots, \theta_K^I$ depending on the state of the latent process I_t (Peña and Guttman, 1988):

$$\begin{aligned}x_t &= F_t x_{t-1} + u_t + G_t w_t, & w_t &\sim N(0, \text{Diag}(\sigma_1^2, \dots, \sigma_d^2)), \\y_t &= H_t x_t + v_t, & v_t &\sim N(0, \theta_{I_t}^I).\end{aligned}$$

Alternatively, one or all of the variances $\sigma_1^2, \dots, \sigma_d^2$ of the transition equation may be Markov switching heteroskedastic (Kim, 1993; Engle and Kim, 1999). Another way of including a switching mechanism is to assume that a drift term is present in the transition equation which switches between various values $\theta_1^I, \dots, \theta_K^I$ (Kim, 1994). For an application of the related switching dynamic factor model to modelling business cycles see Kim and Nelson (1998) and Kaufmann (1999); an interesting application of a state space model with two switching mechanism to audio and signal processing appears in Godsill (1997).

Estimation of a switching state space model is far from trivial. The classical maximum likelihood approach can not be applied directly, as the marginal likelihood where both latent processes

*Department of Statistics, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: sfruehwi@isis.wu-wien.ac.at

x^N and I^N are integrated out, is not available in a closed form. Approximate filters which lead to an unknown approximation error have been used e.g. in Kim (1993, 1994). Alternatively, a Bayesian approach can be applied by means of Markov chain Monte Carlo (MCMC) methods (see e.g. Smith and Roberts, 1993 for a general introduction to MCMC methods). The design of suitable MCMC methods to estimate a Gaussian state space model with switching has been studied in various papers. A general discussion may be found in Carter and Kohn (1994, 1996), Shephard (1994), and Kim and Nelson (1999). MCMC sampling for specific Gaussian state space models with switching appear e.g. in Godsill (1997), Kim and Nelson (1998), Engle and Kim (1999), and Kaufmann (1999). In section 2 of the present paper we address once more the issue of Bayesian estimation of switching Gaussian state space models via MCMC methods. Emphasis will lie on the unidentifiability of switching state space models and its implication for MCMC estimation. We will discuss the problem of label switching, the importance of finding a sensible identifiability constraint, and the fact, that important information such as smoothed estimates of the latent state process x^N or estimates of time varying parameters will be available without the need to identify the model.

In section 3 we head for a fully Bayesian analysis of switching Gaussian state space models discussing issues in model selection such as selecting the number of states and testing for the presence of a hidden Markov switching process. We will apply the method of bridge sampling to compute the model likelihood for a switching Gaussian state space model.

Applications to simulated data appear in section 4. The paper is concluded with an application of a switching state space model to analyse the U.S./U.K. real exchange rate in section 5.

2 Bayesian Estimation of a Gaussian State Space Model with Switching

Unknown parameters which have to be estimated from the data are the fixed parameters θ^C , which are common to all states, the state specific parameters $\theta_1^I, \dots, \theta_K^I$, and the parameter η appearing in the definition of the distribution of $I^N = (I_0, \dots, I_N)$. These parameters will be summarized by ϕ : $\phi = (\theta^C, \theta_1^I, \dots, \theta_K^I, \eta)$. Within the Bayesian approach both the discrete, latent process I^N as well as the continuous, latent process x^N are viewed as missing data and estimated along with the model parameter ϕ . In what follows we are going to estimate the augmented parameter vector $\psi = (\phi, I^N, x^N)$ by sampling from the posterior density $\pi(\psi|y^N)$ by means of MCMC methods.

2.1 Model structure and Choice of the Priors

Bayesian estimation of the model is based on the hierarchical structure of the model:

1. Conditionally on known realisations x^N and I^N of the continuous state process and the switching process, respectively, and on a known model parameter ϕ the observations are independent with the distribution of y_t depending on x_t and I_t , only. The "complete data likelihood" $f(y^N|x^N, I^N, \phi)$ given x^N , I^N and ϕ reads

$$f(y^N|x^N, I^N, \phi) = \prod_{t=1}^N f_N(y_t|x_t, \theta_{I_t}^I, \theta^C),$$

where $f_N(y_t|\cdot)$ is the density a normal distribution.

2. Conditionally on a known realisation I^N of the switching process and on a known model parameter ϕ the density of the (prior) distribution of the continuous latent process $x^N = (x_0, x_1, \dots, x_N)$ is given by:

$$\pi(x^N|I^N, \phi) = \prod_{t=1}^N f_N(x_t|x_{t-1}, \theta_{I_t}^I, \theta^C)\pi(x_0|\phi).$$

3. Conditionally on a known model parameter ϕ the density of the (prior) distribution of the latent switching process I^N depends on η , only: $\pi(I^N|\phi) = \pi(I^N|\eta)$.
4. Finally, ϕ has a prior distribution $\pi(\phi)$.

Note that the „prior” on I^N and x^N appearing in the second and the third level are not subjective priors, but part of the model. One possible prior structure on I^N is the *exchangeable prior*, where I_t is assumed to be an iid process with $Pr\{I_t = j\} = \eta_j, j = 1, \dots, K$. This prior has been used in combination with Gaussian state space models e.g. in Shumway and Stoffer (1991). An alternative choice is the *Markovian switching prior* where I_t is assumed to be a stationary Markov process with discrete state space $\{1, \dots, K\}$ and $Pr\{I_t = j | I_{t-1} = i\} = \eta_{ij}$. This structure has been introduced by Hamilton (1989) and has since been widely used in the time series literature. It is combined with Gaussian state space models e.g. in Kim (1994), Engle and Kim (1999) and Kim and Nelson (1998, 1999).

Only the prior on ϕ appearing on the fourth level has a subjective flavour. From a theoretical point of view, being fully non-informative about ϕ is possible only for the state independent parameter θ^C . Being fully non-informative about $\theta_1^I, \dots, \theta_K^I$ and η is not possible. For switching models it is common to assume that $\eta = (\eta_1, \dots, \eta_K)$, where $\eta_i = (\eta_{i1}, \dots, \eta_{iK})$, is independent from the remaining parameters of ϕ and that all conditional transition distributions $\eta_i, i = 1, \dots, K$, are independent *a priori* from each other. A „natural” prior distribution $\pi(\eta_i)$ for η_i is the Dirichlet prior which is the conjugate prior in the complete data setting, where I^N is assumed to be known. Concerning the state specific parameters $\theta_1^I, \dots, \theta_K^I$, we assume that they are independent *a priori* and that each θ_j^I has the same prior distribution depending on hyperparameters which are not state specific. This allows different parameters for the various states, however with a slight restriction expressed by the prior. Furthermore this prior is invariant to relabeling the states.

2.2 Bayesian Posterior Analysis and the Problem of Label Switching

From the hierarchical structure of the model we obtain that the (unconstrained) posterior density $\pi(\psi|y^N)$ is given by:

$$\pi(\psi|y^N) \propto f(y^N|x^N, I^N, \phi)\pi(x^N|I^N, \phi)\pi(I^N|\phi)\pi(\phi).$$

For models including a latent, discrete structure such as I^N the unconstrained posterior has some characteristic properties (see Stephens, 1997; Celeux, 1998; Frühwirth-Schnatter, 1998). The unconstrained parameter space contains $K!$ subspaces, each one corresponding to a different way of labeling the states. The „complete data likelihood” $f(y^N|x^N, I^N, \phi)$, and the „priors” $\pi(x^N|I^N, \phi)$ and $\pi(I^N|\phi)$ are invariant to relabeling the states. Therefore, if the prior $\pi(\phi)$ is invariant, too, the unconstrained posterior typically is multimodal and invariant to relabeling the states. This special structure of the posterior has a lot of consequences for estimation. The unconstrained model is not identified in a strict sense. When sampling from the unconstrained posterior via MCMC methods, we do not know to which of the labeling subspaces the sampled value belongs. Therefore, we do not know to which state a sampled parameter belongs as label switching (jumping between the various labeling subspaces) might have occurred. Thus we are not allowed to estimate functionals $f(\psi)$ of ψ which are not invariant to relabeling the states from MCMC simulations from the unconstrained posterior. Typical examples are estimation of the state specific parameter such as $\theta_1^I, \dots, \theta_K^I$ and η or estimating the probability $Pr(I_t = j|y^N)$ of being in state j at time t .

2.3 Estimation without Introducing Unique Labelling

Note, that important information is available from MCMC simulations from the unconstrained posterior *without* introducing unique labeling, as we are allowed to estimate functionals $f(\psi)$ of ψ which are invariant to relabeling the states of I_t . This is important especially for applied time series modelling, as „smoothing” in terms of estimating the latent continuous process x^N from

the posterior $\pi(x^N|y^N)$ is possible without caring about identifiability. Smoothed estimates $\hat{x}_{t|N}$ of x_t , for instance, are simply given by:

$$\hat{x}_{t|N} = \frac{1}{M} \sum_{m=1}^M (x^N)^{(m)},$$

where $(x^N)^{(1)}, \dots, (x^N)^{(M)}$ are MCMC simulations from the unconstrained posterior. This estimate is unaffected by label switching, because the marginal posterior $\pi(x^N|y^N)$ is invariant to relabeling the states of I^N .

An other important application is estimation of time varying parameters. In practice, the discrete values $\theta_1^I, \dots, \theta_K^I$ often will be just an approximation to modelling the time varying nature of a certain model parameters ξ_t of the state space model (Harrison and Stevens, 1976):

$$\xi_t = \theta_j^I, \quad \text{if } I_t = j.$$

From rewriting this functional in the form

$$\xi_t = \sum_{j=1}^K \theta_j^I S_t^{(j)},$$

where for each j $S_t^{(j)} \equiv 1$, iff $I_t = j$ and zero otherwise, invariance to relabeling the states of I_t is obvious. Therefore it is possible to obtain individual estimates of ξ_t for each t from the MCMC output of an unconstrained model by averaging $(\theta_s^I)^{(m)}$, where $s = I_t^{(m)}$, over all $m = 1, \dots, M$.

Finally, estimation of state independent parameters θ^C is possible without carrying about identifiability.

2.4 Estimation from an Identified Models

In order to estimate functionals $f(\psi)$ of ψ which are not invariant to relabeling the states such as $\theta_1^I, \dots, \theta_K^I$, η or $Pr(I_t = j|y^N)$ we have to identify the model in the sense that we allow for MCMC simulations from a unique labeling subspace, only.

A common way to do this is to include an identifiability constraint. However, the problem with identifiability constraints is they not necessarily induce a unique labeling, if the geometry of the unconstrained posterior density is ignored (see Frühwirth-Schnatter, 1998). Only a carefully selected constraint will separate the labeling subspaces and induce unique labeling. We will demonstrate in our case studies, how suitable identifiability constraints may be found by exploring MCMC simulations from unconstrained posterior densities.

2.5 MCMC Methods

The design of suitable MCMC methods to generate a (dependent) sample $\psi^{(1)}, \psi^{(2)}, \dots$ from the posterior $\pi(\psi|y^N)$ of a dynamic linear model with switching has been studied in various papers (Carter and Kohn, 1994, 1996; Shephard, 1994; Godsill, 1997; Kim and Nelson, 1998; Engle and Kim, 1999). MCMC techniques for sampling from a complicated posterior density split the joint unknown parameter into blocks and sample then from the conditional posterior densities of each block given the fixed values for the other blocks. Sampling ψ from the posterior of a switching Gaussian state space model is possible within the following four blocks:

- (i) Sample I^N from $\pi(I^N|x^N, \eta, \theta^C, \theta^I, y^N)$
- (ii) Sample η from $\pi(\eta|I^N)$
- (iii) Sample $(\theta_1^I, \dots, \theta_K^I, \theta^C)$ from $\pi(\theta_1^I, \dots, \theta_K^I, \theta^C|x^N, I^N, y^N)$.
- (iv) Sample x^N from $\pi(x^N|I^N, \theta_1^I, \dots, \theta_K^I, y^N)$

Step (i) and (ii) are standard steps occurring for MCMC estimation of any model including a latent Markov switching variable. Step (i) may be carried out in a multimove manner as in Carter and Kohn (1994, 1996), Shephard (1994), or Chib (1996). Sampling η is completely standard, as the conditional posterior $\pi(\eta|\phi, I^N, x^N, y^N)$ depends only on I^N and each conditional distribution η_1, \dots, η_K follows a Dirichlet distribution. The precise procedure applied within step (iii) depends on the specific Gaussian state space model under consideration. For many important examples of state space models such as the basic structural model, the dynamic trend model (Harvey, 1989) or for regression models with random coefficients, H_t , F_t , G_t and u_t are predetermined, and the variances are the only model parameters. For such a conditionally heteroscedastic Gaussian state space model the variances are conditionally independent inverted gamma random variables and easy to sample. Further blocking is necessary, if F_t or H_t depend on unknown model parameters. Finally, step (iv) may be carried out by one of the multimove methods discussed in Carter and Kohn (1994), Frühwirth-Schnatter (1994) and DeJong and Shephard (1995).

There exist various ways to run through this scheme and the suitable method for MCMC sampling depends on what kind of inference is of interest. An unconstrained model may be estimated by *unconstrained Gibbs sampling* running through step (i)-(iv) without any constraint on the state specific parameters. Unconstrained Gibbs sampling, however, does not explore the whole unrestricted parameter space, but tends to stick at the current labeling subspace with occasionally switching to other labeling subspaces, while others will never be visited. An alternative method of estimating an unconstrained model is *random permutation sampling* (Frühwirth-Schnatter, 1998). This method is simply an unconstrained Gibbs sampler concluded by a randomly selected permutation $\rho(1), \dots, \rho(K)$ of the current labeling $1, \dots, K$. After sampling ψ by an unconstrained Gibbs sampler, state dependent parameters are permuted in the following way:

$$(1) \quad \begin{aligned} (\theta_1^I, \dots, \theta_K^I) &:= (\theta_{\rho(1)}^I, \dots, \theta_{\rho(K)}^I), \\ (\eta_{i1}, \dots, \eta_{iK}) &:= (\eta_{\rho(i), \rho(1)}, \dots, \eta_{\rho(i), \rho(K)}), \quad i = 1, \dots, K, \\ (I_1, \dots, I_N) &:= (\rho(I_1), \dots, \rho(I_N)), \end{aligned}$$

whereas the state independent parameters θ^C and x^N remain unchanged. The permuted parameters are the starting point for the next Gibbs step. This sampler is an appropriate method for exploring the whole space of the unconstrained posterior. It delivers a sample from the unconstrained posterior where balanced label switching occurs as all labeling subspaces are visited with the same probability.

To estimate a constrained model an identifiability constraint may be introduced into the sampling scheme. One way of forcing the identifiability constraint is to introduce some truncation or rejection method into step (iii) in order to obtain simulations which fulfill the constraint. This *constrained Gibbs sampling* is the standard method applied so far (see e.g. Engle and Kim, 1999). As mentioned before, identifiability constraints do not necessarily induce a unique labeling, and therefore, constrained sampling may introduce a bias towards a poor constraint. The pooriness of the constraint may go undetected, if we use constrained Gibbs sampling and the sampler sticks at the current labeling subspace.

An alternative method for constrained sampling is *permutation sampling under an identifiability constraint* (Frühwirth-Schnatter, 1998). Unconstrained Gibbs sampling is concluded by a permutation as in (1), but this time the permutation is selected in such a way that the identifiability constraint is fulfilled. This method delivers a sample from the constrained posterior. If the constraint is poor in the sense that it does not induce a unique labeling, the permutation sampler will indicate this fact and exhibit label switching. In this case more suitable identifiability constraint may be found by exploring the MCMC output of random permutation sampling.

3 Issues in Model Selection

3.1 The Bayesian Approach towards Model Selection

In practical time series analysis we may end up with various dynamic linear models $\mathcal{M}_1, \dots, \mathcal{M}_L$ – with or without switching – as possible explanation of our observations. In Frühwirth-Schnatter (1995) it has been suggested to compare state space models by their posterior probabilities $P(\mathcal{M}_l | y^N) \propto L(y^N | \mathcal{M}_l)P(\mathcal{M}_l)$, where $L(y^N | \mathcal{M}_l)$ is the model likelihood obtained by integrating the complete data likelihood $f(y^N | \psi)$ with respect to the prior of all unknown quantities including the latent processes:

$$(2) \quad L(y^N | \mathcal{M}_l) = \int f(y^N | \psi) \pi(\psi) \nu(d\psi).$$

This Bayesian approach towards model selection could be applied to the following issues:

- *Selecting the number of states.* We may compare various switching Gaussian state space models differing in the number K of states.
- *Testing for the presence of Markov switching heterogeneity.* We may compare a non-switching state space model ($K = 1$) with various switching state space models.
- *Testing for constancy of continuous, latent components.* We may compare a model where a certain variance σ_j^2 of the transition equation is positive with a model where this variance is equal to zero.

Most of these test problems are non-regular problems. Testing for the presence of Markov switching heterogeneity and selecting the number of states, for instance, is not possible within the classical framework of maximum likelihood. Although a non-switching state space model could be viewed as that special case of a switching state space model where $K = 1$, the regularity conditions for justifying the χ^2 -approximation to the likelihood ratio statistic do not hold, as the state dependent parameters are unidentified under the hypothesis that there is really one state.

The Bayesian approach compares all possible models under consideration through their model likelihoods $L(y^N | \mathcal{M}_l)$. The computation of the model likelihood which by definition is the normalizing constant of the non-normalized posterior $f(y^N | \psi) \pi(\psi)$, however, has proven to be extremely challenging especially for models with various sets of latent processes such as the switching Gaussian state space model. Analytical integration of (2) with respect to the whole parameter ψ is not possible and we have to apply some approximation method. Model likelihoods have been estimated from the MCMC output using methods such as the candidate's formula (Chib, 1995), importance sampling based on mixture approximations (Frühwirth-Schnatter, 1995), combining MCMC simulations and asymptotic approximations (Gelfand and Dey, 1994; DiCiccio *et al.*, 1997) and bridge sampling (Meng and Wong, 1996).

Computing the model likelihood from the MCMC output of a switching model without a latent continuous state process x^N has been discussed in detail in Frühwirth-Schnatter (1999) with the following main results. First, estimation of the model likelihood turns out to be sensitive to the problem of label switching. Especially the candidate's formula (Chib, 1995) should not be applied, if label switching is present, and is therefore suitable only for models with unique labeling. Second, the best result with the lowest standard error is obtained by using the method of bridge sampling (Meng and Wong, 1996). In the present paper we extend this method to Gaussian state space models with switching. We will give details in the next subsection.

3.2 Computing the Model Likelihood from the MCMC Output

For a model including both discrete as well as continuous latent variables I^N and x^N , respectively, there are basically two ways of reducing the dimension of integration in (2). If conditionally on I^N and ϕ the state space model is a linear Gaussian state space model, then it is possible to integrate out the continuous latent process x^N by running Kalman filtering conditional on I^N

and ϕ . The second method – which does not rely on Gaussianity nor linearity of the state space model – is to integrate out the switching process I^N :

$$(3) \quad L(y^N) = \int L(y^N | x^N, \phi) \pi(x^N | \phi) \pi(\phi) d(x^N, \phi),$$

where closed formulae for the marginal likelihood $L(y^N | x^N, \phi)$ and the marginal prior $\pi(x^N | \phi)$ are available.

Subsequently, we will discuss evaluation of the second integral (3) from the MCMC output using the bridge sampling technique (Meng and Wong, 1996). Bridge sampling is a method for computing ratios of normalising constant from MCMC simulations of the posterior. DiCiccio *et al.* (1997) suggested to apply bridge sampling to the problem of computing the model likelihood. It is obvious from (3) that the model likelihood is equal to the normalising constant of the posterior density $\pi(x^N, \phi | y^N) \propto \pi^*(x^N, \phi | y^N) = L(y^N | x^N, \phi) \pi(x^N | \phi) \pi(\phi)$. Let $q(x^N, \phi)$ be a density with known normalising constant, which is some simple approximation to the posterior $\pi(x^N, \phi | y^N)$. Let $\alpha(x^N, \phi)$ be an arbitrary function such that $\int \alpha(x^N, \phi) \pi(x^N, \phi | y^N) q(x^N, \phi) d(x^N, \phi) > 0$. Bridge sampling is based on the following result:

$$1 = \frac{\int \alpha(x^N, \phi) \pi(x^N, \phi | y^N) q(x^N, \phi) d(x^N, \phi)}{\int \alpha(x^N, \phi) q(x^N, \phi) \pi(x^N, \phi | y^N) d(x^N, \phi)} = \frac{\int \alpha(x^N, \phi) \pi^*(x^N, \phi | y^N) q(x^N, \phi) d(x^N, \phi)}{L(y^N) \int \alpha(x^N, \phi) q(x^N, \phi) \pi(x^N, \phi | y^N) d(x^N, \phi)},$$

which yields the key identity:

$$(4) \quad L(y^N) = \frac{E_q(\alpha(x^N, \phi) \pi^*(x^N, \phi | y^N))}{E_\pi(\alpha(x^N, \phi) q(x^N, \phi))},$$

where E_f is the expectation with respect to a density $f(\cdot)$. If – dependent or independent – samples $(x^N, \phi)^{(m)}$, $m = 1, \dots, M$, and $(\tilde{x}^N, \tilde{\phi})^{(l)}$, $l = 1, \dots, L$ from the posterior $\pi(x^N, \phi | y^N)$ and the approximate density $q(x^N, \phi)$, respectively, are available, then both expectations are substituted by the appropriate averages and we obtain the bridge sampling estimator $\hat{L}_{BS}(y^N)$:

$$(5) \quad \hat{L}_{BS}(y^N) = \frac{\hat{E}_q}{\hat{E}_\pi} = \frac{L^{-1} \sum_{l=1}^L \alpha((\tilde{x}^N, \tilde{\phi})^{(l)}) \pi^*((\tilde{x}^N, \tilde{\phi})^{(l)} | y^N)}{M^{-1} \sum_{m=1}^M \alpha((x^N, \phi)^{(m)}) q((x^N, \phi)^{(m)})}.$$

Note that there are two functions for tuning: the bridge function $\alpha(x^N, \phi)$ and the importance density $q(x^N, \phi)$.

If one uses $\alpha(x^N, \phi) = 1/q(x^N, \phi)$, for instance, then we obtain

$$L(y^N) = E_q \left(\frac{\pi^*(x^N, \phi | y^N)}{q(x^N, \phi)} \right)$$

and the following importance sampling estimator of the model likelihood of a switching state space model:

$$(6) \quad \hat{L}_{IS}(y^N) = \frac{1}{L} \sum_{l=1}^L \frac{\pi^*((\tilde{x}^N, \tilde{\phi})^{(l)} | y^N)}{q((\tilde{x}^N, \tilde{\phi})^{(l)})}$$

with $(\tilde{x}^N, \tilde{\phi})^{(1)}, \dots, (\tilde{x}^N, \tilde{\phi})^{(L)}$ being an i.i.d. sample from $q(x^N, \phi)$. This estimator is an extension of the importance sampling estimator of Frühwirth-Schnatter (1995) to Gaussian state space models with switching.

If one uses $\alpha(x^N, \phi) = 1/\pi^*(x^N, \phi | y^N)$, the basic identity for reciprocal importance sampling (Gelfand and Dey, 1994) results:

$$L(y^N) = \left\{ E_\pi \left(\frac{q(x^N, \phi)}{\pi^*(x^N, \phi | y^N)} \right) \right\}^{-1},$$

leading to the following reciprocal importance sampling estimator of the model likelihood of a switching state space model: :

$$(7) \quad \hat{L}_{RI}(y^N) = \left\{ M^{-1} \sum_{m=1}^M \frac{q((x^N, \phi)^{(m)})}{\pi^*((x^N, \phi)^{(m)} | y^N)} \right\}^{-1}.$$

Meng and Wong (1996) discuss an asymptotically optimal choice of $\alpha(x^N, \phi)$, which minimizes the expected relative error of the estimator $\hat{L}_{BS}(y^N)$ for iid draws from $\pi(x^N, \phi|y^N)$ and $q(x^N, \phi)$:

$$(8) \quad \alpha(x^N, \phi) \propto \frac{1}{L \cdot q(x^N, \phi) + M \cdot \pi(x^N, \phi|y^N)}.$$

We will refer to the corresponding bridge sampling estimator as the „optimal” bridge sampling estimator. As the optimal choice depends on the *normalized* posterior, we apply the following iterative procedure: based on a previous estimate $\hat{L}_{BS}^{(t-1)}(y^N)$ of the normalizing constant, the posterior is normalized, $\hat{\pi}(x^N, \phi|y^N) = \pi^*(x^N, \phi|y^N) / \hat{L}_{BS}^{(t-1)}(y^N)$, and a new estimate $\hat{L}_{BS}^{(t)}(y^N)$ is computed by (5). This leads to the following recursion:

$$(9) \quad \hat{L}_{BS}^{(t)}(y^N) = \hat{L}_{BS}^{(t-1)}(y^N) \frac{L^{-1} \sum_{l=1}^L \frac{\hat{\pi}((\tilde{x}^N, \tilde{\phi})^{(l)}|y^N)}{L \cdot q((\tilde{x}^N, \tilde{\phi})^{(l)}) + M \cdot \hat{\pi}((\tilde{x}^N, \tilde{\phi})^{(l)}|y^N)}}{M^{-1} \sum_{m=1}^M \frac{q((x^N, \phi)^{(m)})}{L \cdot q((x^N, \phi)^{(m)}) + M \cdot \hat{\pi}((x^N, \phi)^{(m)}|y^N)}}.$$

Either the importance sampling estimator or the reciprocal importance sampling estimator may serve as starting value $\hat{L}_{BS}^{(0)}(y^N)$.

In Frühwirth-Schnatter (1999), the „optimal” bridge sampling estimator outperformed both importance sampling as well as reciprocal importance sampling for the following reason: whereas importance sampling as well as reciprocal importance sampling are known to be sensitive to the tail behaviour of the importance density $q(\cdot)$, it has been shown in Frühwirth-Schnatter (1999) that the „optimal” bridge sampling estimator is much more robust in this concern. This is of special importance for switching state space models, where an importance density in the space of the unobserved, latent process x^N has to be constructed. The choice of an importance density which has only fat or only thin tails in *all* directions of the parameter space appears unattainable and robustness to the tail behaviour will be of great importance.

Now we turn to the choice of $q(x^N, \phi)$ for switching state space models. For much simpler models than a switching Gaussian state space model, DiCiccio *et al* (1997) suggested to construct a *normal* importance density from the MCMC output, e.g. by fitting a quadratic function to the log of non-normalised marginal posterior. As already pointed out by DiCiccio *et al.* (1997) this choice will work only for well-behaved, regular problems, where the posterior is close to a normal density. One has, however, to be careful with normal approximations to the unconstrained posterior density $\pi(x^N, \phi|y^N)$ for switching state space models. For an unconstrained model the posterior usually is multimodal with more or less well-separated modes and a single normal importance density might be an extremely bad choice. Therefore we construct the importance density in an unsupervised manner from MCMC simulations $\psi^{(1)}, \dots, \psi^{(M_L)}$ from the unconstrained posterior $\pi(\psi|y^N)$ in a similar way as in Frühwirth-Schnatter (1995, 1999):

$$(10) \quad q(x^N, \phi) = \frac{1}{M_L} \sum_{n=1}^{M_L} \pi(\eta|(I^N)^{(n)}) \pi(x^N|\theta^{(n)}, (I^N)^{(n)}, y^N) K_\theta(\theta|(I^N, x^N, \theta)^{(n)}, y^N),$$

where $K_\theta(\theta|I^N, x^N, \theta', y^N)$ is the density of the transition kernel appearing in the unconstrained Gibbs sampler. In what follows we assume that the normalising constant of this density is available. The mixture importance density (10) automatically deals with the problem of multimodality. The mixture approximation is based on averaging over the conditional densities, where the argument (x^N, ϕ) is fixed and the conditioning indicator process I^N is sampled from the unconstrained posterior switching between the different ways of labeling. Therefore the mixture importance density (10) will be multimodal, too. In order to reproduce all modes of the posterior, however, it is essential to base the mixture approximation (10) on a MCMC method which forces *balanced* label switching such as the random permutation sampler.

4 Application to Simulated Data

4.1 Data simulated from Model with Switching

First, we apply the methods of the previous sections to data simulated from a local level model with switching observation variance:

$$\begin{aligned}x_t &= x_{t-1} + w_t, & w_t &\sim N(0, Q_t), \\y_t &= x_t + v_t, & v_t &\sim N(0, R_t),\end{aligned}$$

where Q_t is constant ($Q_t = 0.001$), R_t is a switching variance:

$$R_t = \begin{cases} 0.01, & I_t = 1, \\ 0.1, & I_t = 2, \end{cases}$$

and I_t is a 2-state Markov chain with transition matrix

$$\eta = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}.$$

We test the „true” model against the following alternatives:

- a local level model with jointly Markov switching variances („switching model 1”):

$$(Q_t, R_t) = \begin{cases} (Q^{[1]}, R^{[1]}), & I_t = 1, \\ (Q^{[2]}, R^{[2]}), & I_t = 2; \end{cases}$$

- a local level model with two independently Markov switching variances („switching model 2”):

$$Q_t = \begin{cases} Q^{[1]}, & I_t^1 = 1, \\ Q^{[2]}, & I_t^1 = 2, \end{cases} \quad R_t = \begin{cases} R^{[1]}, & I_t^2 = 1, \\ R^{[2]}, & I_t^2 = 2; \end{cases}$$

an extension of this model, where R_t switches between three states:

$$R_t = \begin{cases} R^{[1]}, & I_t^2 = 1, \\ R^{[2]}, & I_t^2 = 2, \\ R^{[3]}, & I_t^2 = 3; \end{cases}$$

- and a local level model without switching variances.

In figure 1 we perform explorative Bayesian analysis of the MCMC output obtained from random permutation sampling of these models. For a model with a single switching variable all simulations $(Q^{[j]}, R^{[j]})^{(l)}$, $l = 1, \dots, L$, are projected onto the (Q, R) -plane for all states $j = 1, \dots, K$. Similarly, for all models with two switching variables all simulations $(Q^{[j]}, R^{[j']})^{(l)}$, $l = 1, \dots, L$, are projected onto the (Q, R) -plane for all combinations of states $j \in \{1, \dots, K_1\}$ of the first and states $j' \in \{1, \dots, K_2\}$ of the second switching variable.

There are a lot of interesting hints in these figures concerning the model selection issues discussed earlier. For the model with just one switching variable („switching model 1”) there are two groups, as expected, and although the model is the wrong one, the simulations show that there is practically no difference between Q in the first and the second state. This finding would lead to hypothesize the correct model. If we carry out formal model selection between „switching model 1” and the true model (table 1), we find that the true model is clearly preferred. For the model with two switching variables („switching model 2”) again there are just two groups both for two and three states in the observation variance R . This is not what would be expected, if the selected number of states were correct. If we had two states in both variances, we would expect four groups in this figure and similarly, six groups, if one variance has two and the other variance has three different states. The fact that there are just two groups is a hint that the number of states is too large for both models. The MCMC output of both models clearly

indicates, that there are two states in R and just one state in Q . Again we would hypothesize the true model from analyzing the MCMC output of a model which is wrong. If we test all wrong models against the true one (table 1), we find that the true model has the biggest model likelihood. For completeness, we also report the model likelihood for a model without switching which is much smaller than the model likelihood for any of the switching models.

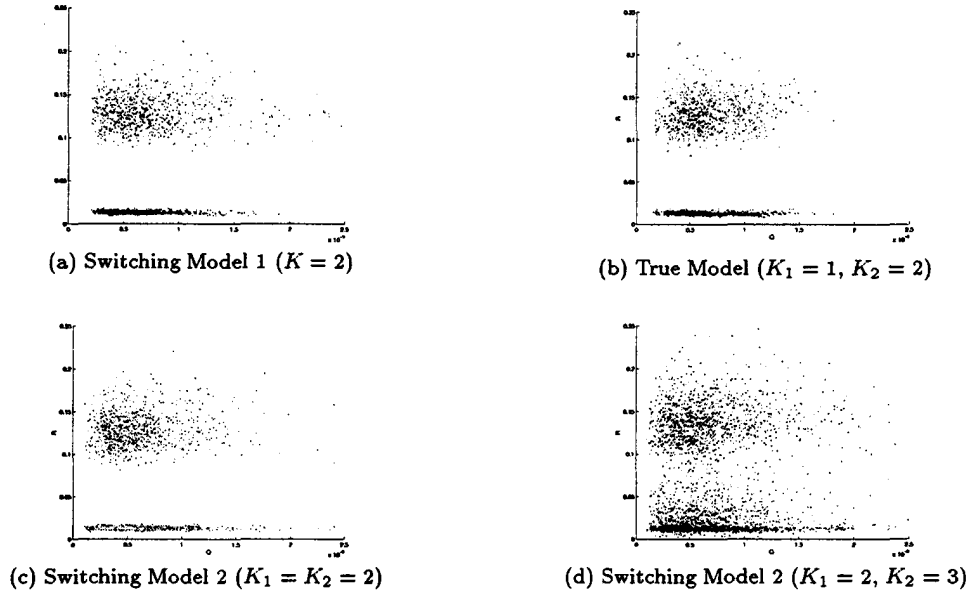


Figure 1: Explorative Bayesian analysis for simulated data set 1

Table 1: Formal model selection for simulated data set 1 (standard errors given in parenthesis)

Model	$\log L(y^N \text{Model})$
True Model ($K_1 = 1, K_2 = 2$)	57.3052 (0.0819)
Switching Model 1 ($K = 2$)	49.6716 (0.0665)
Switching Model 2 ($K_1 = 2, K_2 = 2$)	45.8414 (0.0845)
Switching Model 2 ($K_1 = 2, K_2 = 3$)	33.9585 (0.0928)
No switching	-6.2878 (0.0427)

After having selected the model, we need an identifiability constraint, if we are interested in state specific estimation. From figure 1 it is clear that the constraint $R^{[1]} < R^{[2]}$ separates the groups. Including this constraint into the constrained permutation sampler leads to an identified model without label switching.

4.2 Data simulated from a Model without Switching

Next, we simulated data from a local level model without switching:

$$\begin{aligned} x_t &= x_{t-1} + w_t, & w_t &\sim N(0, Q_t) \\ y_t &= x_t + v_t, & v_t &\sim N(0, R_t), \end{aligned}$$

where Q_t and R_t are constant: $Q_t \equiv 0.001$, $R_t \equiv 0.01$. We test this model against the following alternatives:

- a local level model with jointly switching variances („switching model 1”):

$$(Q_t, R_t) = \begin{cases} (Q^{[1]}, R^{[1]}), & I_t = 1, \\ (Q^{[2]}, R^{[2]}), & I_t = 2; \end{cases}$$

- and a local level model with two independently switching variances („switching model 2”):

$$Q_t = \begin{cases} Q^{[1]}, & I_t^1 = 1, \\ Q^{[2]}, & I_t^1 = 2, \end{cases} \quad R_t = \begin{cases} R^{[1]}, & I_t^2 = 1, \\ R^{[2]}, & I_t^2 = 2. \end{cases}$$

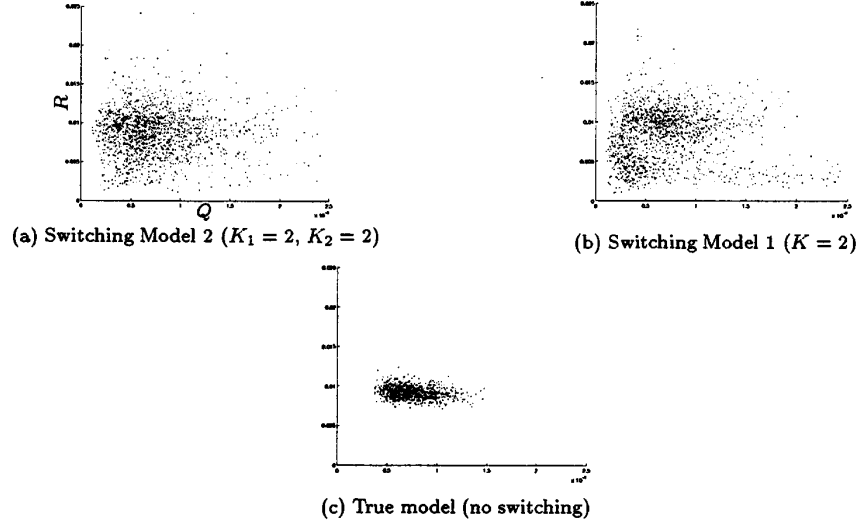


Figure 2: Explorative Bayesian analysis for simulated data set 2

In figure 2 we perform explorative Bayesian analysis of the MCMC output obtained from random permutation sampling of these models, where the simulations are projected in the same way as in figure 1. The simulation from „switching model 2” should show four groups, if both variances were switching independently, however the only group we see is a clear hint that this model should be compared with a model without switching which would be the true one. „Switching model 1” shows two groups which are not very well separated. If we compute model likelihoods for all models (table 2), we find that the true model, the one without switching, has the highest model likelihood.

Table 2: Formal model selection for simulated data set 2 (standard errors given in parenthesis)

Model	$\log L(y^N \text{Model})$
True Model (No switching)	299.0033 (0.0251)
Switching Model 1 ($K = 2$)	282.9084 (0.0879)
Switching Model 2 ($K_1 = 2, K_2 = 2$)	280.8277 (0.0817)

5 Application to Modelling Exchange Rate Data

For further illustration we reanalyze the U.S./U.K. real exchange rate from January 1885 to November 1995, originally published in Grilli and Kaminsky (1991) and reanalyzed by Engle and Kim (1999). The real exchange rate is defined as the relative price of U.K. to U.S. producer goods, i.e. U.S./U.K. nominal exchange rate times the U.K. producer price index divided by the U.S. producer price index. Engle and Kim (1999) suggested to decompose the log of the real exchange rate y_t into a permanent component p_t and a transitory component c_t :

$$\log y_t = p_t + c_t,$$

where p_t follows a random walk process,

$$p_t = p_{t-1} + w_{2,t}, \quad w_{2,t} \sim N(0, \sigma_{2,t}^2),$$

whereas c_t is assumed to follow an AR(r)-process:

$$c_t = \phi_1 c_{t-1} + \dots + \phi_r c_{t-r} + w_{t,2}, \quad w_{t,2} \sim N(0, \sigma_{1,t}^2).$$

The variance $\sigma_{1,t}^2$ of the transitory component c_t is assumed to switch between K_1 states $\theta_1^{1,I}, \dots, \theta_{K_1}^{1,I}$ according to a Markov switching process I_t^1 with transition matrix η^1 , whereas the variance $\sigma_{2,t}^2$ of the permanent component is assumed to switch between K_2 states $\theta_1^{2,I}, \dots, \theta_{K_2}^{2,I}$ according to a Markov switching process I_t^2 with transition matrix η^2 :

$$(11) \quad \sigma_{1,t}^2 = \theta_{I_t^1}^{1,I}, \quad \sigma_{2,t}^2 = \theta_{I_t^2}^{2,I}.$$

As the model can be reformulated as a dynamic linear model with state vector $x_t = (p_t, c_t, \dots, c_{t-r})$, it can be viewed as a special case of the switching Gaussian state space model discussed in the present paper. Various issues in model selection will be discussed within this case study:

- Is the (conditional) variance of the permanent and/or the transitory component really Markov switching?
- How many states do the switching variances exhibit?
- How should we select the order r of the AR(r)-process?

In their paper Engle and Kim (1999) selected a model where the (conditional) variance of the transitory component is determined from a 3-state Markov switching process, the (conditional) variance of permanent component is constant, and the order of the AR-process is equal to two (i.e. $K_1 = 3, K_2 = 1, r = 2$). They adopt this specification by exploring the posterior distributions without formal Bayesian model selection. In the present paper, however, we end up with a different specification as we start with explorative Bayesian analysis of a more general model with $K_1 = 4, K_2 = 2$ and $r = 3$. The best model will be one where the (conditional) variance of the transitory component is determined from a four-state Markov switching process, the (conditional) variance of permanent component is constant, and the order of the AR-process is equal to two (i.e. $K_1 = 4, K_2 = 1, r = 2$). Our findings will be confirmed by formal Bayesian model selection.

Also, our estimation scheme differs from the one adopted in Engle and Kim (1999). First, we do not condition on the first values of the state process, but sample the whole processes $c_{1-r}, \dots, c_0, \dots, c_N$ and p_0, \dots, p_N including the starting values by applying the multi-move sampler of Frühwirth-Schnatter (1994), and initializing the filtering step with the prior $x_0 \sim N(\hat{x}_{0|0}, P_{0|0})$, where $\hat{x}_{0|0} = (100 \ 0 \ \dots \ 0)'$ and $P_{0|0} = \text{Diag}(1000 \ 0 \ \dots \ 0)$. Sampling the AR(r)-parameters is carried out in a similar fashion as in Engle and Kim (1999) from the regression model $c_t = c_{t-1}\phi_1 + \dots + c_{t-r}\phi_r + (\theta_{I_t^1}^{1,I})^{0.5} \cdot \varepsilon_t$, where ε_t is iid standard normal. However, as samples of c_0, \dots, c_{1-r} are available from our Gibbs sampler, t is running from 1 to N rather than from $r+2$ to N . Within one iteration, sampling ϕ_1, \dots, ϕ_r is repeated until the stationarity condition on the AR(r)-process is fulfilled.

Second, we start with estimating unidentified models of various orders and try to find the „best“ one through formal Bayesian model selection. Estimation is based on random permutation sampling rather than the method of single move sampling from truncated densities discussed in Engle and Kim (1999). Within permutation sampling we may sample all variances $\theta_i^{1,I}$, $i = 1, \dots, K_1$, and $\theta_j^{2,I}$, $j = 1, \dots, K_2$, at the same time, as they are conditionally independent, inverted Gamma distributed for an unconstrained model given $(I^{1,N}, I^{2,N}, x^N, y^N)$. Estimation is based on the symmetric priors $\theta_i^{1,I} \sim \text{IG}(3, 8)$, $i = 1, \dots, K_1$, and $\theta_j^{2,I} \sim \text{IG}(3, 2)$, $j = 1, \dots, K_2$. The prior for all conditional transition probabilities η_i^1 , and η_j^2 is chosen to be $D(1, \dots, 1)$.

We start with explorative Bayesian analysis of a model with $K_1 = 4, K_2 = 2$ and $r = 3$. We used the random permutation sampler for estimation, where each sampling step is concluded by a randomly selected permutation. Although this model is not identified, a lot of interesting information is available from the MCMC output. Part (a) and (b) of figure 3, for instance, show a scatter plot of (stationary) MCMC simulations $(\theta_i^{1,I})^{(m)}$ versus $(\eta_{ii}^1)^{(m)}$ and $(\theta_j^{2,I})^{(m)}$

versus $(\eta_{jj}^2)^{(m)}$ for all possible states $i \in \{1, \dots, K_1\}$ and $j \in \{1, \dots, K_2\}$, respectively. For I_t^1 we allowed for four states and there are actually four distinct groups, for I_t^2 , however, we allowed for two states and there are no separate groups. This provides empirical evidence in favour of a homogeneous rather than a switching variance of the permanent component. This hypothesis is further supported by part (c) of the figure where $(\theta_i^{1,I})^{(m)}$ is plotted versus $(\theta_j^{2,I})^{(m)}$ for all possible combinations (i, j) of states. Finally, part (d) of the same figure plots the posterior of the AR-parameter ϕ_3 which may be estimated directly from the output of the random permutation sampler as ϕ_3 is state independent. The mode of the posterior is close to 0 providing evidence for the hypothesis that ϕ_3 is equal to zero. To sum up, from our explorative analysis we obtain evidence in favour of a model with $K_1 = 4$, $K_2 = 1$ and $r = 2$ rather than $K_1 = 4$, $K_2 = 2$ and $r = 3$.

In table 3 both models are compared by formal Bayesian model selection, showing a much higher model likelihood for the simpler model. Explorative Bayesian analysis of the new model is reported in figure 4. The scatter plot of $(\theta_i^{1,I})^{(m)}$ versus $(\eta_{ii}^1)^{(m)}$ for all possible states $i \in \{1, \dots, K_1\}$ still shows four groups. The posterior of ϕ_2 is shifted away from 0, but there is some evidence for the hypothesis that ϕ_2 actually is 0. The model likelihood of the model with $K_1 = 4$, $K_2 = 1$ and $r = 1$ reported in table 3, however, clearly favours the model with $K_1 = 4$, $K_2 = 1$ and $r = 2$ which is our final choice. This model differs from the one selected in Engle and Kim (1999) by the number of states of the variance of the transitory component. The model likelihoods in table 3 clearly favour our „best” model rather than the one selected by Engle and Kim (1999). Increasing the number of states from four to five, however, reduces the model likelihood drastically. For completeness, we report the model likelihood for a model without switching, showing that this model is the most unlikely of all. Computation of all model likelihoods reported in table 3 is based on the random permutation sampler.

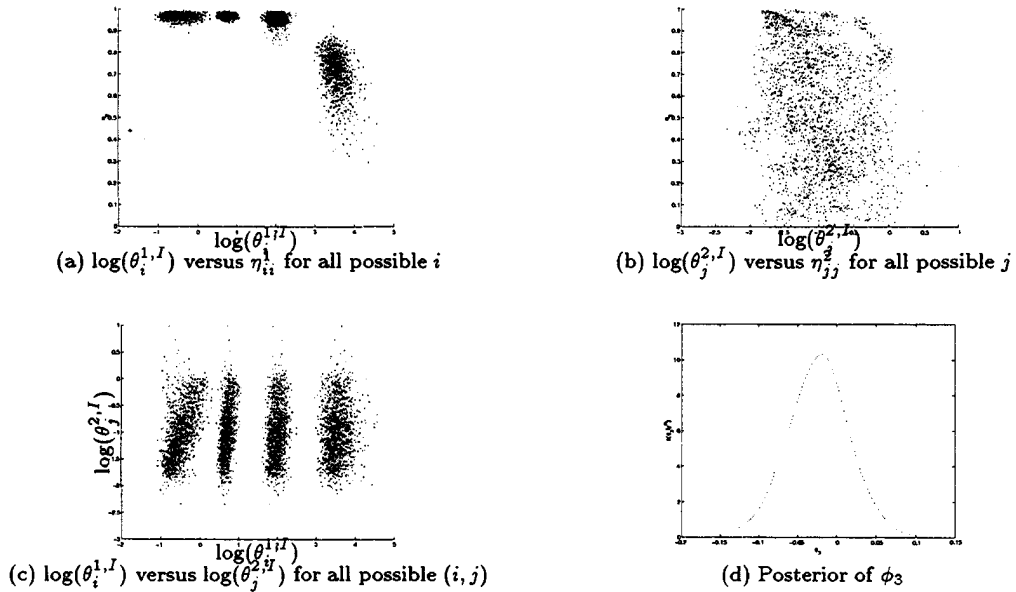


Figure 3: Explorative Bayesian analysis for a switching model with $K_1 = 4$, $K_2 = 2$, $r = 3$

We can draw further interesting inference from the output of the random permutation sampler without the need to identify the model. This is especially true for the smoothed permanent component which is compared in figure 5 with the observed time series. Our estimate of the permanent component is much smoother than the rather noisy estimate published in Engle and Kim (1999), being nearly constant till the end of the fifties and increasing afterwards. Another interesting picture is obtained, if we plot the time varying variance $\sigma_{1,t}^2$ estimated from (11) by

$$\hat{\sigma}_{1,t}^2 = \frac{1}{M} \sum_{m=1}^M (\theta_t^{1,I})^{(m)},$$

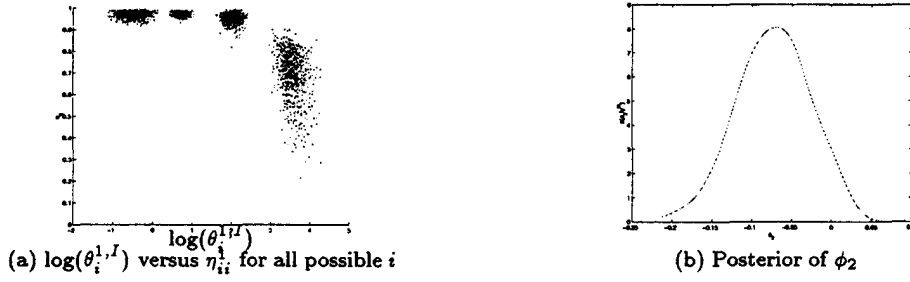


Figure 4: Explorative Bayesian analysis for a switching model with $K_1 = 4, K_2 = 1, r = 2$

Table 3: Formal model selection for exchange rate data
(Standard error given in parenthesis)

Model	$\log L(y^N \text{Model})$
$K_1 = 4, K_2 = 2, r = 3$	-2562.4 (0.217)
$K_1 = 4, K_2 = 1, r = 2$	-2515.5 (0.083)
$K_1 = 4, K_2 = 1, r = 1$	-2612.5 (0.124)
$K_1 = 3, K_2 = 1, r = 2$	-2605.9 (0.076)
$K_1 = 5, K_2 = 1, r = 2$	-2880.2 (0.120)
No switching, $r = 2$	-2914.4 (0.081)

where $s = (I_t^1)^{(m)}$ over time t (figure 6). Again no identification was necessary to obtain this picture.

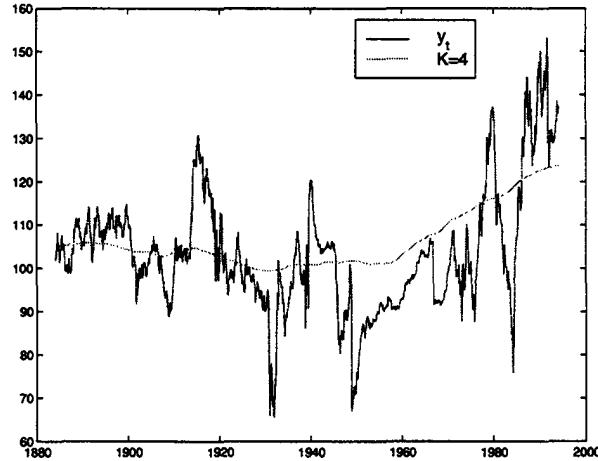


Figure 5: Smoothed Real Exchange Rate for a four state model ($K_1 = 4, K_2 = 1, r = 2$)

To draw state specific inference on the variances of the different states as well as to obtain the smoothed posterior state probabilities, however, we have to identify the model. For comparison, we will identify both the „best” model and the three state model selected by Engle and Kim (1999). An identifiability constraint for the „best” model is obvious from figure 4. The main difference between the states is the associated variance of the transitory component leading to the constraint $\theta_1^{1,I} < \theta_2^{1,I} < \theta_3^{1,I} < \theta_4^{1,I}$. If this constraint is included into the permutation sampler under identifiability constraints, no label switching occurs. Table 4 reports point estimates as well as 95%-H.P.D.-regions for all model parameters, including estimates of the state specific variances as well as estimates of the transition probabilities. Similarly, the identifiability constraint for the three state model is given by: $\theta_1^{1,I} < \theta_2^{1,I} < \theta_3^{1,I}$.

Figure 7 plots the estimated smoothed posterior state probabilities $Pr(I_t^1 = i | y^N)$ of being in a certain state $i \in \{1, 2, 3, 4\}$ over time t , for a four state switching model and compares

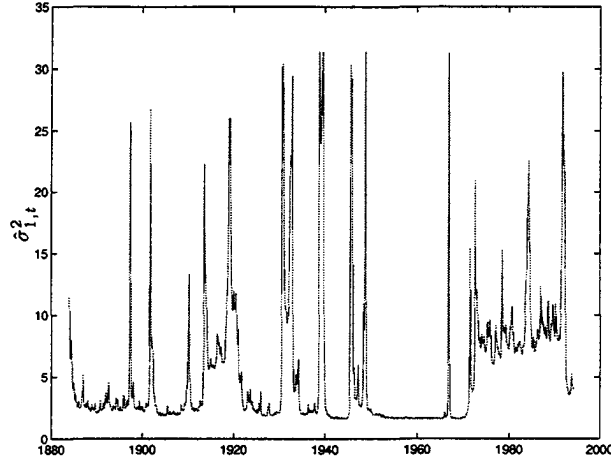


Figure 6: Estimated time varying variance $\hat{\sigma}_{1,t}^2$ for a four state model ($K_1 = 4$, $K_2 = 1$, $r = 2$)

them with the probabilities obtained from the three state model. The probabilities $Pr(I_t^1 = i|y^N)$, $i = 1, \dots, K_1$ are estimated from the constrained MCMC output by:

$$\hat{Pr}(I_t^1 = i|y^N) = \frac{1}{M} \# \{ (I_t^1)^{(m)} = i \}.$$

For the three state model Engle and Kim (1999) found that periods with low-state variance correspond to periods in which the nominal exchange rate was fixed, whereas periods with medium-state variance correspond to periods of floating nominal exchange rates. Periods of high-state variance are rather singular events and can be identified with specific historical events. It is interesting to observe that for the four state model the third and the fourth state actually correspond to the second and the third state of the three state model, whereas the first two states are a further split of the low variance state of the three state model. The most quiet state occurred during the first half of the forties and then from about 1952 to the end of the seventies.

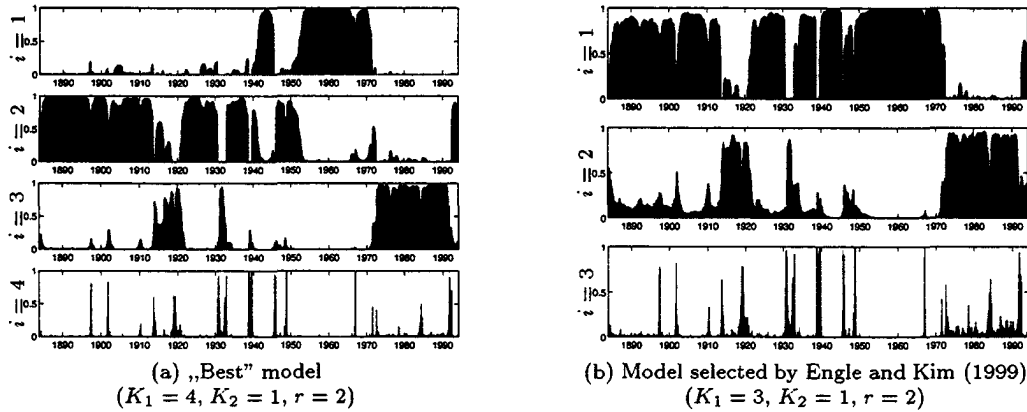


Figure 7: Smoothed Probabilities

References

Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika*, **81**, 541-53.

Table 4: Estimation Results for $K_1 = 4, K_2 = 1, r = 2$

Parameter	Mean	Std.dev.	95%-H.P.D. regions	
θ_1^I	0.634	0.151	0.371	0.93
θ_2^I	2.05	0.196	1.67	2.42
θ_3^I	7.63	1.07	5.9	9.88
θ_4^I	36.4	9.13	20.7	53.9
σ_2^2	0.366	0.132	0.121	0.608
ϕ_1	1.06	0.0474	0.967	1.14
ϕ_2	-0.0729	0.046	-0.158	0.0139
η_{11}	0.968	0.0132	0.943	0.991
η_{12}	0.0091	0.00861	2.84e-006	0.0256
η_{13}	0.00639	0.00586	2.87e-006	0.0189
η_{14}	0.0162	0.00987	0.000231	0.0341
η_{21}	0.00855	0.00576	0.000165	0.0205
η_{22}	0.973	0.00853	0.957	0.988
η_{23}	0.00587	0.0057	6.19e-006	0.0155
η_{24}	0.0123	0.00697	0.000484	0.0246
η_{31}	0.00498	0.00489	1.24e-005	0.0144
η_{32}	0.0139	0.0123	9.59e-006	0.0373
η_{33}	0.956	0.0222	0.916	0.992
η_{34}	0.0248	0.0161	0.00129	0.0562
η_{41}	0.039	0.0338	0.000159	0.103
η_{42}	0.147	0.0691	0.024	0.288
η_{43}	0.123	0.0934	0.00108	0.309
η_{44}	0.691	0.116	0.438	0.865

Carter, C.K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models, *Biometrika*, **83**, 589-601.

Chib, S. (1995). Marginal Likelihoods from the Gibbs Output, *Journal of the American Statistical Association*, **90**, 1313-21.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics*, **75**, 79-97.

Celeux, G. (1998). *Bayesian Inference for Mixture: The Label Switching Problem*, Preprint Inria.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models, *Biometrika*, **82**, 339-350.

DiCiccio, T.J., Kass, R., Raftery, A. and L. Wasserman (1997). Computing Bayes Factors By Combining Simulations and Asymptotic Approximations, *Journal of the American Statistical Association*, **92**, 903-15.

Engle, C. and Kim, C.-J. (1999). The long-run U.S./U.K. real exchange rate, *Journal of Money, Credit, and Banking*, **31**, 335-356.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, **15**, 183-202.

Frühwirth-Schnatter, S. (1995). Bayesian Model Discrimination and Bayes Factors for Linear Gaussian State Space Models, *Journal of the Royal Statistical Society, B* **57**, 237-246.

Frühwirth-Schnatter, S. (1998). MCMC Estimation of Classical and Dynamic Switching and Mixture Models, Technical report 1998-61, Department of Statistics, Vienna University of Economics and Business Administration.

Frühwirth-Schnatter, S. (1999). Model Likelihoods and Bayes Factors for Switching and Mixture Models, Technical report 1999-71, Department of Statistics, Vienna University of Economics and Business Administration.

- Gelfand, A.E., and D.K. Dey (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society*, **B 56**, 501-514.
- Godsill, S.J. (1997). Bayesian enhancement of speech and audio signals in the presence of both impulsive and background noise. *International Statistical Review*, **65**, 1-21.
- Grilli, V. and Kaminsky, G. (1991). Nominal exchange rate regimes and the real exchange rate: Evidence from the United States and Great Britain, 1885-1986, *Journal of Monetary Economics*, **27**, 191-212.
- Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357-84.
- Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting (with discussion), *Journal of the Royal Statistical Society*, **B 38**, 205-247.
- Harvey, A.C (1989). *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.
- Kaufmann, S. (1999). Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods. Working paper, revised September 1999, Department of Economics, University of Vienna.
- Kim, C.-J. (1993). Unobserved-Component Time Series Models with Markov-Switching Heteroscedasticity. *Journal of Business & Economic Statistics*, **11**, 341-349.
- Kim, C.-J. (1994). Dynamic linear models with Markov Switching, *Journal of Econometrics*, **60**, 1-22.
- Kim, C.-J. and Ch.R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching. *Review of Economics and Statistics*, **80**, 188-201.
- Kim, C.-J. and Ch.R. Nelson (1999). *State-Space Models with Regime Switching*. MIT Press, Cambridge MA.
- Meng, X.L. and W.H. Wong (1996). Simulating Ratios of Normalising Constants via a Simple Identity. *Statistica Sinica*, **6**, 831-60.
- Peña, D. and Guttman, I. (1988). Bayesian approach to robustifying the Kalman filter, *Bayesian Analysis of Time Series and Dynamic Linear Models* (ed. J.C. Spall), 227-254, Marcel Dekker, New York.
- Shephard, N. (1994). Partial Non-Gaussian state space models, *Biometrika*, **81**, 115-31.
- Shumway, R.H. and Stoffer, D.S. (1991). Dynamic linear models with switching, *Journal of the American Statistical Association*, **86**, 763-9.
- Smith, A.F.M. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, *Journal of the Royal Statistical Society*, **B 55**, 3-23.
- Stephens, M. (1997). *Bayesian Methods for Mixtures of Normal Distributions*. Ph.D.Thesis, University of Oxford.